Further Observations on the Role of Silence in the Perception of Stop Consonants*

M. F. Dorman,† L. J. Raphael,‖ and A. M. Liberman‖‖

ABSTRACT

Previous research has shown that silence is a necessary condition for the perception of stop consonants in such utterances as /sa/ vs. /sta/. In this case, there are cues in the vocalic portion of the syllable that are appropriate for stop-consonant manner. Hence, the silence was shown only to be a necessary, not a sufficient, cue for stop manner. In the present series of experiments we have investigated the role of silence in the perception of stops when there are no stop manner cues in the vocalic portion. The purpose was to determine the extent to which the silence cue can be not only a necessary, but also a sufficient cue.

The experiments reported here were intended to shed light on the role of silence in the perception of prevocalic stop and affricate consonants. More specifically, we were concerned with the following questions: (1) Is silence a sufficient cue to the perception of these consonants?, and (2) Is the perceptual processing of silence phonetic as well as auditory?

Previously reported experiments (Raphael, Dorman, and Liberman, 1976) have indicated that silence is a necessary but not sufficient cue to the perception of stop manner in prevocalic, as well as other, positions; that is, for example, in order to hear the stop in a contrast such as [sɛ] vs. [spɛ], we need the s-friction, a silent interval, and a vocalic segment that, in isolation, is heard as [pɛ]. Thus, silence, in this situation, may simply allow time for the [pɛ] to evade masking by the friction. But if [spɛ] could be produced, similarly, from sound segments that contain no stop—that is, if silence could be shown to be a sufficient cue—then a masking interpretation would be ruled out.

In the situations described so far, silence is seen to be a necessary but not sufficient cue for the perception of a stop manner; that is, in order

*This is a slightly revised version of a paper presented at the 91st meeting of the Acoustical Society of America, Washington, D.C., 4-9 April 1976.

†Also Arizona State University, Tempe, Arizona.

‖Also Herbert H. Lehman College and the Graduate School of the City University of New York.

‖‖Also University of Connecticut, Storrs.

[HASKINS LABORATORIES: Status Report on Speech Research SR-48 (1976)]
to hear the stop in the [s]-[sp] contrast, for example, we need the s-friction, a silent interval, and a vocalic segment that, in isolation, is heard as [pc]. Thus, silence, in this situation, may simply allow time for the [pc] to evade masking by the friction. But if [sp] could be produced, similarly, from sound segments that contain no stop—that is, if silence could be shown to be a sufficient cue—then a masking interpretation would be ruled out.

It is of interest, then, to ask whether there are situations in which silence is, in fact, a sufficient cue for perception of stop manner. We have long suspected that there are, given the finding many years ago that [slit] can be changed to [split] by inserting an appropriate amount of silence between the [s] friction and the onset of the vocalic segment (Liberman, Harris, Eimas, Lisker, and Bastian, 1961). That experiment was done by cutting apart and reassembling the frictional and vocalic portions of the syllable [slit] and recorded on magnetic tape; hence there may have been transients at the onset of the vocalic portion that were more or less sufficient, given the necessary amount of silence preceding them, to produce perception of the stop. We have, therefore, repeated and extended that experiment, but in a way more suitable to our purposes. To avoid having stop-relevant cues in either the frictional or vocalic components, we recorded, separately, the hissing noise of [s] and the syllable [lit]. Having determined then that there was no perceptible [p] in the noise or in the syllable [lit], we used the pulse-code-modulation (PCM) system at Haskins Laboratories to put the noise before the [lit] with various intervals of silence in between. We then randomized the resulting patterns, and presented them to listeners for judgment. The results are shown in Figure 1. We see that at silent intervals less than 70 msec, listeners reported [slit]—that is, they did not hear a stop—but when the silent interval was greater than 70 msec, they reported [split]—that is, they heard the stop consonant. We may conclude, then, that in this case silence is a sufficient cue for perception of the stop manner.

How do we interpret the role of silence in the experiment just described? It seems to us plain that the silence provides information, not time to evade masking. The information is that the speaker either did or did not close his vocal tract appropriately for production of the [p] segment in the syllable [split]. But, to use that information in order to arrive at the perception of a stop would appear to require a process that is not merely auditory but is, rather, more abstractly phonetic.

If the role of silence is, in fact, to provide information about vocal-tract closure necessary for the production of a stop, we might suppose that a stop would not be heard at silent intervals greater than those produced by normal articulation. To test that expectation, we carried out a second experiment in which we extended the silent intervals of the first experiment all the way out to 650 msec. The outcome is shown in Figure 2. We see there that, as in the first experiment, [slit] is heard at relatively short intervals of silence, and [split] at somewhat longer intervals. We also see that at still longer intervals of silence the listeners once again hear [slit]. Thus, intervals of silence much longer than those that characterize the stop closure do not produce the perception of the stop.

Let us turn now to another phonetic contrast for which the presence or absence of the stop closure is again a distinguishing articulatory feature,
Figure 1: Percent identification of "split" as a function of the interval between /s/ and /lɪt/.
Figure 2: Percent identification of "split" as a function of the interval between /s/ and /lnt/.
but which differs from [slit] vs. [split] in two respects: the stop closure occurs between syllables, not within a syllable, and the contrast is between fricative and affricative. The example is the difference between the utterances "Please say shop" and "Please say chop." Is silence a sufficient cue here, too? To find out, we performed the following experiment.

Having recorded the utterance "Please say shop," we separated "please say" from "shop" and then recombined them with silent intervals between "say" and "shop" that varied from 0 to 150 msec. The resulting patterns were randomized and presented to listeners with instructions to identify each one as "please say shop" or "please say chop." The results are shown in Figure 3. We see that at intervals of less than about 50 msec, listeners reported hearing "please say shop," while at longer intervals they heard "please say chop." Thus, silence is a sufficient condition for the perception of the affricate in absolute initial position in the syllable.

Figure 3: Percent identification of "chop" as a function of the interval between the carrier phrase "please say" and "shop."

We suppose that here, too, silence provides the (phonetic) information that the speaker either did or did not close his vocal tract in the manner necessary for the production of the affricate. And here, too, we thought it
at silent intervals of 0 to 100 msec between the end of "say" and the beginning of "shop." These patterns were randomized and presented to listeners for judgment as "please say shop" or "please say chop."

The results are shown in Figure 5. The solid curve is the labeling function for the slow condition. We find the boundary at about 23 msec. The boundary for the fast condition, represented by the dashed line, lies at about 30 msec. Thus, there is an effect of rate, but its direction indicates that at the faster speech rate our listeners needed a longer silent interval before they perceived the affricate in "shop." At the present time we cannot interpret that finding with confidence, since, as we said, we have not measured the effects of speaking rate on articulation of the various parts of the utterance. We can only remind ourselves that the duration of the friction is also a cue for the distinction between fricative and affricative—longer for fricative, shorter for affricative—and then suppose that, given the fast precursive phrase, the duration of the friction would seem longer and hence would bias the perception toward the fricative. Conceivably, that overrode the contrary bias that might have been expected from the silence cue. In any case, we may conclude that the effect of the silence cue is, in fact, sensitive to speaking rate. To interpret that result, we shall have to invoke processes that go beyond the psychophysics of gap perception.

We should summarize. We have found that silence can be a sufficient cue for the perception of the stop (in [slit] vs. [split]) and the affricate (in [shop] vs. [chop]): Moreover, the effective cue is not just any duration of silence, nor is it precisely that duration which is perceptibly different from no silence; it is, rather, a range of durations that appear to match reasonably well the silences that result from the vocal tract closures when the stop and affricate are produced. In the case of the affricate/fricative distinction, the effect of the silence cue was also found to be sensitive to the rate of articulation. We suppose that the role of silence is to provide information about the presence or absence of the vocal-tract closure that marks the production of the stop and affricate phones. Accordingly, we assume that the perceptual processing of the silence cue is not only auditory but also phonetic.

REFERENCES


Summerfield, A. Q. (1975b) How a full account of segmental perception